ETL-0561

Built-Up Area Feature Extraction
First Year Report

David M. McKeown, Jr.
Aviad Zlotnick

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, Pennsylvania 15213

February 1990

Approved for public release; distribution is unlimited.

Prepared for:

U.S. Army Corps of Engineers
Engineer Topographic Laboratories
Fort Belvoir, Virginia 22060-5546

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
February 1990

3. REPORT TYPE AND DATES COVERED
Annual FROM 87/7 TO 88/6

4. TITLE AND SUBTITLE
Built-Up Area Feature Extraction: First Year Report

5. FUNDING NUMBERS
DACA72-87-C-0001

6. AUTHOR(S)
McKeown, David M. Jr. and Zlotnick, Aviad

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Carnegie-Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Topographic Laboratories
CEETL-RI-T
Fort Belvoir, VA 22060-5546

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ETL-0561

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release;
Distribution is Unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report describes research performed by the Digital Mapping laboratory at Carnegie-Mellon University on the analysis of aerial images of builtup areas during the first year of Contract DACA72-87-C-0001. This research can be divided into three major parts: (1) extracting road networks from images; (2) detecting and delineating buildings; and (3) basic research to support extraction of the above features.

Previous work in large-scale spatial databases and in knowledge-based systems for scene interpretation is described. Research results performed under this Built-Up Area Feature Extraction contract are described. New research in road network extraction is discussed. New work in the use of structural analysis to hypothesize and verify buildings using monocular cues in complex imagery is also discussed. Finally, the current state of research including successes, failures, and goals for the second year continuation are described.

14. SUBJECT TERMS
Feature extraction — Road networks
Scene interpretation — Built-up area

15. NUMBER OF PAGES
39

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	UNCLASSIFIED	UL

Preface

This report describes work performed under contract DACA72-87-C-0001, by Carnegie Mellon University, Pittsburgh, Pennsylvania, for the U.S. Army Engineer Topographic Laboratories (ETL), Fort Belvoir, Virginia. The Contracting Officer's Representative at ETL is Mr. George E. Lukes.

Table of Contents

List of Figures

List of Tables

1. Introduction

In July, 1987 the Digital Mapping Laboratory at Carnegie-Mellon University started work on the analysis of aerial images of builtup areas, supported by USAETL contract DACA72-87-C-0001. Our research in the course of the first year, ending in June '88, can be divided into three major parts:

1. Extracting road networks from images.
2. Detecting and delineating buildings.
3. Basic research to support extraction of the above features.

In the following Section we describe, as background, some previous work in large-scale spatial databases and in knowledge-based systems for scene interpretation. Sections 3 and 4 describe research results performed under this Built-Up Area Feature Extraction contract. In Section 3 we discuss our new research in road network extraction. In Section 4 we describe new work in the use of structural analysis to hypothesize and verify buildings using monocular cues in complex aerial imagery. Finally, in Section 5 the current state of our research program including successes, failures, and goals for the second year continuation.

2. Background and Previous Work

Over the last eight years our research interests have been to investigate the use of knowledge intensive techniques for the detailed analysis of remotely sensed imagery by developing scene interpretation systems for airports and urban areas. This work has resulted in the design and implementation of several rule-based image analysis systems and supporting work in knowledge acquisition and performance analysis tools. Within the Digital Mapping project, a group of 4 researchers and 6 part-time student programmers, we have developed system implementations for road network extraction, suburban house scene analysis, and airport scene interpretation. Our accomplishments include the following major system implementations:

- Developed MAPS, a large scale image/map database system which integrates digital imagery, terrain data, and map data. A hierarchical representation for flexible spatial queries, and a general image-to-map correspondence allow users to utilize the display of high-resolution imagery to efficiently index into a geodetic referenced map database. Spatial entities in this database are represented as objects with flexible attributes and user-defined relationships.

- Developed the first rule-based scene interpretation system, SPAM, which relies on spatial, spectral, and map knowledge implemented as over 600 production rules. The system domain of expertise is a collection of commercial and military airports. Image analysis tools such as stereo verification, linear alignment, and map-guided region growing were integrated into the rule-based system. We have extended the scope of the SPAM knowledge-based system for aerial image analysis to interpret suburban house scenes from its original task domain of airport scene analysis.

- Within the context of the SPAM system we have pioneered the development of techniques for the automatic compilation of spatial and structural constraints into OPS5 productions, which can be directly executed by the SPAM interpretation system. We have also integrated automated performance evaluation tools in order to evaluate the effect of various types of knowledge in the quality of the overall scene interpretation. This work relies on a database of human generated ground truth interpretations.

- We developed a high performance system for road tracking, ARF, that uses multiple cooperative methods for extracting information about road location and structure from complex aerial imagery. This system is a multi-level architecture for image analysis that allows for cooperation among low-level processes and aggregation of information by high-level analysis components.

In the following Sections 2.1 and 2.2 we provide some background discussion of our work in spatial databases and knowledge based scene interpretation.

2.1. MAPS: Image/Map Spatial Database System

The MAPS spatial database [2, 4, 5, 7] was developed between 1980-1984 supported by the DARPA Image Understanding Program as research into large-scale spatial databases and spatial knowledge representation. It is interesting that this system has expanded from its original research goal of developing an interactive database for answering spatial queries into a component of several knowledge-based image understanding systems under development at Carnegie Mellon University. MAPS is a large-scale image/map database system for the Washington D.C. area that contains approximately 200 high resolution aerial images, a digital terrain database, and a variety of map databases from the United States Geological Survey (USGS) and the Defense Mapping Agency (DMA). MAPS has been used as a component for an automated road finder/follower, a stereo verification module, a 3D scene generation system, and a knowledge-based system for interpreting airport scenes in aerial imager. In addition, MAPS has an interactive user query component that allows users to perform spatial queries using high resolution display of aerial imagery as an method for indexing into the spatial database. This capability to relate, over time, imagery at a variety of spatial resolutions to a spatial database forms a basis for a large variety of interpretation and analysis tasks such as change detection, map update, and model-based interpretation.

Figure 2-1 shows the system organization of MAPS. Four databases are maintained within MAPS: a digital terrain database, a map database, a landmark database, and an image database. A fifth database, CONCEPTMAP, consists of a schema-based representation for spatial entities and a set of procedural methods that provide a uniform interface to each of the four component databases for interactive users or application programs. It is this interface that allows us to represent and access image, map, terrain, and collateral data in a manner that best suits the intrinsic structure of the data. At the same time the CONCEPTMAP database provides uniform access to a variety of spatial data independent of the particular internal structure. This is in sharp

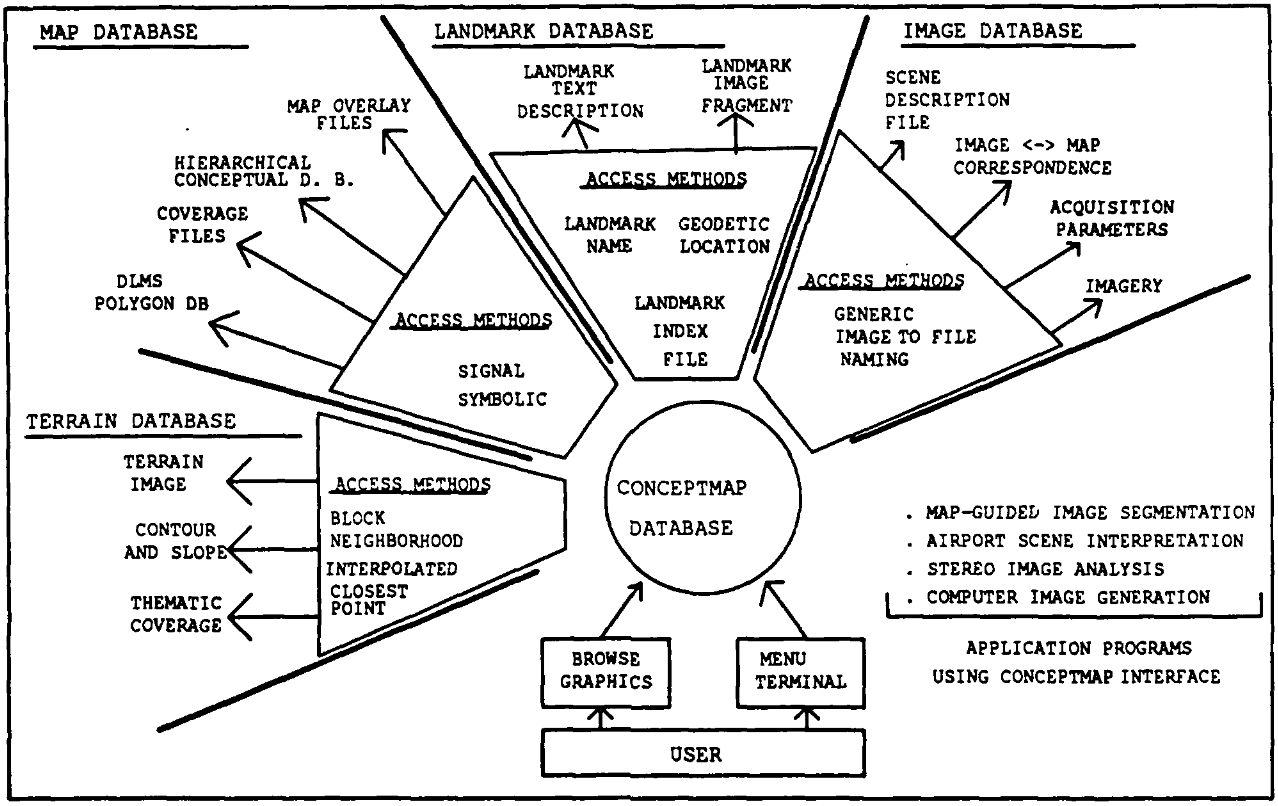

Figure 2-1: MAPS: System Overview

contrast to methods proposed for uniform representation of image and cultural data such as raster data sets and regular decompositions such as quadtrees or k-d trees.

2.1.1. Schema-Based Representation For Spatial Entities

The CONCEPTMAP database uses a *schema-based* representation for spatial entities. Using schemas (or frames) is a well understood AI methodology for representing knowledge. Such a representation can be combined within several problem-solving methods such as semantic networks, scripts or production systems to construct a problem-solving system [1]. Each entity in the CONCEPTMAP database is represented by one *concept* schema and at least one *role* schema. A *concept* can represent any spatial object and associates a name with a set of attributes stored in the *concept* and *role* schemata. A *concept* such as 'washington d.c' might have multiple *role* schemata defined, each representing a different view of the same spatial area, such as 'political', and 'demographic'. A *concept* such as 'georgetown university' might have multiple *role* schemata defined for each of the campus buildings or areas.

There are three unique identifiers generated by the CONCEPTMAP system which allow for indirect access to additional factual properties of concept or role schemata.

- The *concept-id* is unique across all concepts in all CONCEPTMAP databases. That is, given a concept-id one can uniquely determine the name of the spatial entity.

- The *role-id* uniquely determines a role schema across all CONCEPTMAP databases.
- The *role-geographics-id* uniquely determines a collection of points, lines or polygons in vector notation. Each point is represented as <latitude,longitude,elevation>.

These identifiers are also used to index into other components of the MAPS database. For example, the *concept-id* is used to search for landmark descriptions of measured ground control points used during the calculation of transform functions for image-to-map and map-to-image correspondence. The *role-id* is used as the basic entity when building fast access methods to the spatial data utilizing a hierarchy tree decomposition. The *role-geographics-id* is used to acquire the unique geographic position for a *role schema* as well as for linkage into the MAPS image database and segmentation files generated by human interaction or machine segmentation. There are three reasons for this approach. First, it allows CONCEPTMAP to handle very large databases with a minimal amount of information resident in the application process. The identifiers provide a level of indirection to the actual data, which is stored in a variety of formats and may or may not be present for a large subset of the database. Second, we can achieve a great deal of flexibility and modularity in processes which communicate about spatial entities. Given the name of a CONCEPTMAP database, a *concept-id* or *role-id* uniquely determines the entity in question. This facilitates the construction of application programs with simple query structures, requiring a minimum of communication overhead. Finally, given this decoupling from the CONCEPTMAP database, each of the MAPS component databases, image database, terrain database, landmark database, and map database may be physically resident on a different workstation or mainframe.

2.1.2. A Geodetic Frame Of Reference

An implicit requirement crucial to successful operationalization of spatial knowledge for image analysis is that the metrics used by the analysis system be defined in cartographic coordinates, such as *<latitude/longitude/elevation>*, rather than in an image-based coordinate system. Systems that rely on descriptions such as "the runway has area 12000 pixels" or "suburban homes are between 212 and 345 pixels in area" are useless except for (perhaps) the analysis of one image. Further, spatial analysis based on the semantics of *above*, *below*, *left-of*, *right-of* etc., are also inappropriate for general interpretation systems. To operationalize metric knowledge one must relate the world model to the image under analysis. This should be done through image-to-map correspondence using camera models. Every image in the MAPS database has an associated set of ground control points and scene parameter file, including camera focal length, aircraft elevation and position, and digitization information to establish interior orientation. Therefore, we can directly measure ground distances, areas, absolute compass direction, and recover crude estimates of height using a camera model computed for each image under analysis. The MAPS system allows us to express spatial knowledge in ground based metrics, which in turn allows us to work with imagery and map data at a variety of scales and resolutions.

2.2. SPAM: Knowledge-Based Scene Analysis

SPAM, System for Photo interpretation of Airports using MAPS, is an image-interpretation system. SPAM [3, 6, 8] coordinates and controls image segmentation, segmentation analysis, and the construction of a scene model. It provides several unique capabilities to bring map knowledge and collateral information to bear during all phases of the interpretation. These capabilities include:

- The use of domain-dependent spatial constraints to restrict and refine hypothesis formation during analysis.
- The use of explicit camera models that allow for the projection of map information onto the image.
- The use of image-independent metric models for shape, size, distance, absolute and relative position computation.
- The use of multiple image cues to verify ambiguous segmentations. Stereo pairs or overlapping image sequences can be used to extract information or to detect missing components of the model.

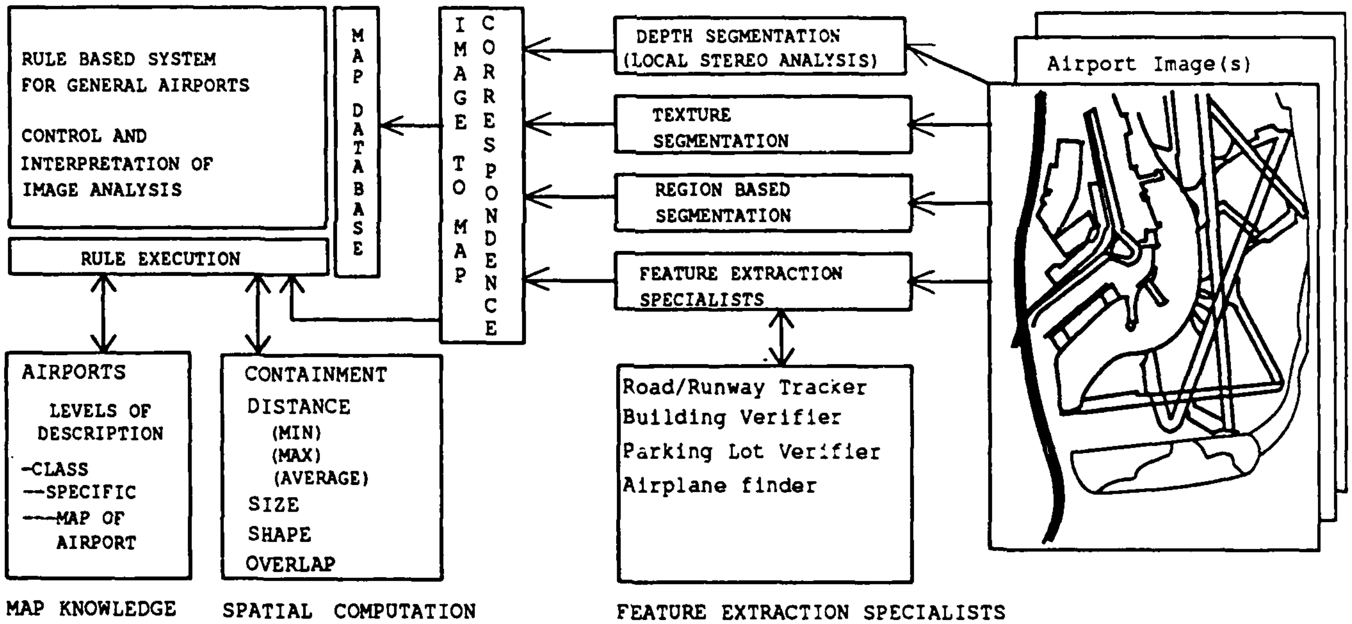

Figure 2-2: SPAM: System Organization

Figure 2-2 shows the overall organization of the interpretation system. SPAM maintains an internal spatial database that is composed of feature extracted from imagery by various methods, possibly from several images, where the features are represented in terms of their geodetic position rather than their image coordinates. In fact, SPAM performs interpretation in map-space which allows for a variety of knowledge such as maps and multi-temporal imagery to be handled in a uniform manner. SPAM uses the following components of the MAPS system.

- CONCEPTMAP to retrieve shape and position models and site-specific map knowledge.
- Map-to-image correspondence to project models onto new imagery and image-to-map

correspondence to calculate metric distances and areas.

- Procedures to compute spatial relationships between hypotheses including containment, intersection, adjacency, closest point of approach, subsumed by.

2.2.1. The SPAM Interpretation Architecture

SPAM represents four types of interpretation primitives, *regions*, *fragments*, *functional areas*, and *models*. SPAM performs scene interpretation by transforming image *regions* into scene *fragment* interpretations, aggregating these fragments into consistent and compatible collections called *functional areas*, and selecting sets of functional areas that form *models* of the scene. Loosely speaking there are four *phases* of interpretation. Each of these four phases operationalizes one or more types of domain knowledge. In order to build a SPAM system we must be able to acquire knowledge for each interpretation phase as described in Figure 2-3. This knowledge is represented as OPS5 production rules. Over 600 productions are currently used in the airport scene interpretation task, nearly half are used to evaluate and propagate local consistency.

Phase 1: Region-to-fragment

Assigns the image region data a set of fragment interpretations based solely on local properties (2-D shape characteristics, texture, 3-D depth/height, etc.) and knowledge about the classes of objects found in the scene.

Phase 2: Local-consistency-check

Pair-wise tests are performed on the fragment interpretations that utilize spatial knowledge about the scene under consideration. The confidence of those interpretations supporting one another are incremented based on the quality of the test.

Phase 3: Functional-area

Sets of mutually consistent interpretations that share similar functions or are spatial decompositions of the scene are grouped into cliques called functional areas.

Phase 4: Model-generation

Sets of functional areas are grouped together into scene segments. The segments with the largest number of functional areas become distinct scene models. Any conflicts encountered when combining functional areas are resolved by a default strategy, using the accumulated support for each interpretation, or by specific knowledge added by the user.

Figure 2-3: Interpretation Phases In SPAM

Each phase is executed in the order given in Figure 2-4. SPAM drives from a local, low-level set of interpretations to a high-level, more global, scene interpretation. There is a set of hard-wired productions for each phase that control the order of rule executions, the forking of processes, and other domain-independent tasks. However this "bottom-up" organization does not preclude interactions between phases. For example, prediction of a fragment interpretation in *functional-area* phase will automatically cause SPAM to reenter *local-consistency* phase for that fragment. Other forms of top-down activity include stereo verification to disambiguate conflicting hypotheses in *model-generation* phase and linear alignment in *region-to-fragment* phase.

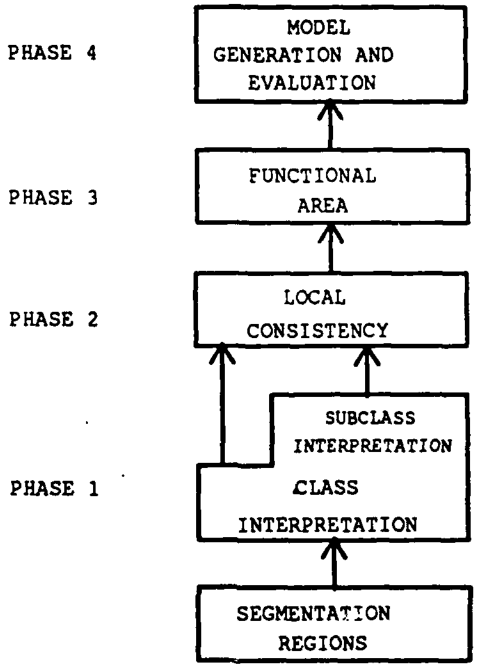

Figure 2-4: Control Flow For Interpretation Phases In SPAM

Figure 2-5 shows the refinement/consistency/prediction paradigm used in SPAM within each interpretation phase. Knowledge is used to check for consistency among hypotheses, to predict missing components using context, and to create contexts based on collections of consistent hypotheses. Prediction is restrained in SPAM in that hypotheses cannot predict missing components at their own representation level. A collection of hypotheses must combine to create a context from which a prediction can be made. These contexts are refinements or spatial aggregations in the scene. For example, a collection of mutually consistent runways and taxiways might combine to generate a runway functional area. Rules that encode knowledge that runway functional areas often contain grassy areas or tarmac may predict that certain sub-areas within that functional area are good candidates for finding such regions. However, an isolated runway or taxiway hypothesis cannot directly make these predictions. In SPAM the context determines the prediction. This serves to decrease the combinatorics of hypothesis generation and to allow the system to focus on those areas with strong support at each level of the interpretation.

2.2.2. The Current State of SPAM

The SPAM system is an ongoing focal point for our research in knowledge-intensive scene analysis. Imagery has been acquired with stereo and multi-temporal coverage for six airports including, Dulles International, National Airport, Andrews AFB, San Francisco International, Los Angeles International, and Moffett Field. The SPAM interpretation system has been run on both machine-generated segmentations as well as human ground truth segmentations. These results have been reported in the literature [3, 6, 8].

The SPAM architecture has also been used to analyze several surburban house scenes containing roads, houses, lawns, and driveways. While this is a much simpler task, it shows that SPAM is capable of performing across different domains, and that task specific knowledge could be decoupled from the overall SPAM interpretation and control strategy.

SPAM requires significant computational resources. It currently takes between 20 and 120 cpu hours on a VAX 11/785 processor to perform a complete run on one of our complex airport scenes. We have begun to investigate opportunities for parallelism within the context of coarse grain task-level decomposition using shared memory multiprocessors such as the Encore MultiMax. In conjunction with the Production Systems Machine (PSM) research project we are utilizing a parallel version of OPS5 and augmenting the run-time environment in order to allow rule actions to proceed in parallel. We see great promise in such architectures.

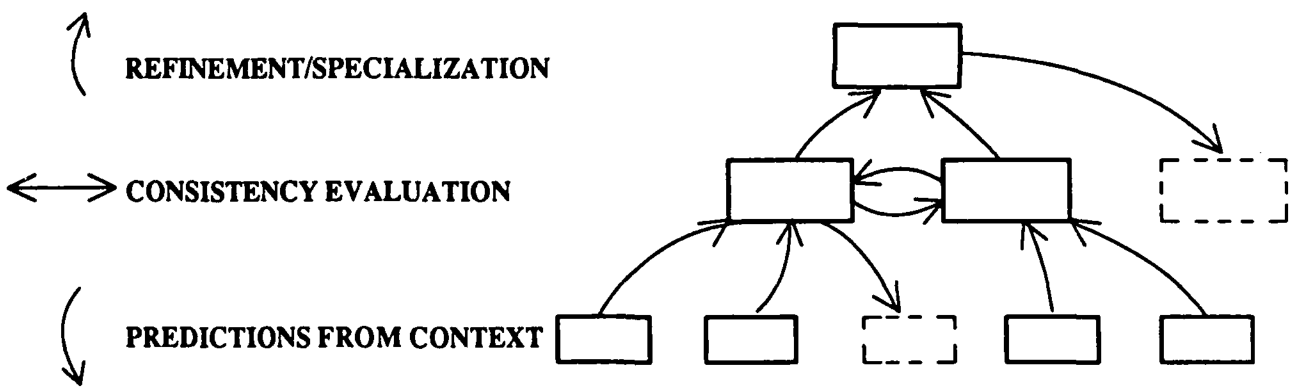

Figure 2-5: Refinement, Consistency, and Prediction in SPAM

A better understanding of how to expand the capability of knowledge-intensive systems such as SPAM requires the exploration of automated techniques for knowledge acquisition and utilization. A fundamental shift in how we engineer such systems is required in order to expand the level of performance and avoid brittle behavior when presented with slightly different task domains. This methodological shift requires basic ideas and theories of how to acquire, integrate, and operationalize large amounts of spatial and structural data into a knowledge base.

3. Road Network Extraction

Roads are one of the most dominant features in aerial imagery, and have much importance in mapping applications. Consequently, there have been many research efforts aimed at extracting roads from remotely sensed imagery. Most of this work equates road finding with linear feature extraction, and can be classified into two categories: road finding and road tracking.

In road *finding* the entire image is processed with an edge operator and anti-parallel edges are selected as roads, implicitely assuming that the road background material is the same on both sides of the road. This approach does capture one of the roads' dominant features, but breaks down when road edges are broken by intersections or when they are fuzzy.

In road *tracking* a segment of a road is assumed to be known, and the image is processed locally around that segment in order to extend it. Local processing has advantages both in speed and in better sensitive to finer details that may be lost in global processing. Most road trackers extend the road based on correlation image intensity profiles of road cross-sections. They usually track well, but are susceptible of locking on to good correlation arising from textured areas that have nothing to do with roads.

The equivalence of road finding with linear feature extraction was justified when only low resolution, such as LANDSAT I,II, was available. It is no longer sufficient, since we now have much higher resolution imagery, and practical mapping applications require detailed structural analysis of the road surface.

The Digital Mapping Project at Carnegie-Mellon University has been involved in creating an integrated road extraction system called RNET. RNET takes an image as input and computes a network of roads and intersections as output. It utilizes the road properties traditionally used for road extraction, namely edges and surface profile, and also geometric knowledge that allows for extending roads over areas where image intensity does not carry enough support for the other criteria. RNET's tracker is unique in combining independent edge-based and correlation-based trackers in a fashion that avoids the weak points of both. RNET also combines road finding and road tracking, enjoying the advantages of both, and breaks away from linear feature detection by delineating road surface changes and overpasses, detecting road width changes and intersections, and computing a symbolic description of the road *network* both as a labeled graph and as a verbal description of the individual roads.

RNET's components are:

1. **The main control loop** determines *areas of interest* and invokes the *road finder* to find *road seeds* in that area.It then selects road seeds and fires the *road tracker* and *network analyzer*. The road seeds are selected so as to minimize multiple tracking of the same roads, with consideration of the fact that road seed ends are sometimes less reliable than the rest of the road seed.

   This program controls the entire road extraction process. Areas of Interest may be either an image section selected by the user, a band around the image boundaries, or the entire image. After invoking the road finder in the Area of Interest, new road seeds are derived from the road finder output, so that there is a road seed inside every finder seed that is long enough. These new road seeds do not suffer from unreliable ends. The control program maintains a bitmap in which tracked roads are marked, and a road seed that is within a marked road is not passed on to the tracker. Derivation of internal road seeds improved tracking considerably, and checking in the bitmap resulted in an order of magnitude speedup without losing roads.

2. **The road finder** scans an area of interest in the image and produces road seeds. changes made since July '87 are:
   a. Improved treatment of road seed splitting during finding: In general, the longer a road seed is, the more confidence we have in its correctness. In the previous road finder version seeds were arbitrarily chopped at splitting points. In the current version competing branches are compared, bad branches are rejected and good ones are kept in their full length, including overlaps. This improvement proved most important where road marking caused fragmentation of road seeds.

      Figure 3-1 shows roads seeds found in a manually selected area of interest. Thin lines are intensity edges and thick lines are detected road seeds. It may be seen that there are some errors in the road finder output. The next paragraph addresses this problem.

   b. Detection and elimination of road linking errors: There is a trade-off between long road seeds and linking errors. Relaxing the thresholds gives longer correct road seeds, but also causes an undesirable increase in the number of errors produced. We have developed a program that detects most of the linking errors. The program is based on an idea published by Martin Gardner in Scientific American. Its use enables us to relax the road finder's thresholds without the penalty of added errors.

      Figure 3-2 shows the effect of removing detected linking errors. The thick lines are the road seeds considered correct, and the thin lines are the detected errors.

   c. Post-linking: After the road finder finishes there still remain roads that are fragmented. The post-linking phase finds pairs of roads that are close and have compatible orientations and links them together. In this way roads seeds that span gaps due to overpasses are generated. The total number of road seeds is reduced, and the reliability of seeds is improved. Post-linking is done by connecting road seeds that are near, and then using the algorithm that detects linking errors.

      Figure 3-3 shows the effect of post-linking. The thick lines are the input road seeds, and the thin lines are the computed links.

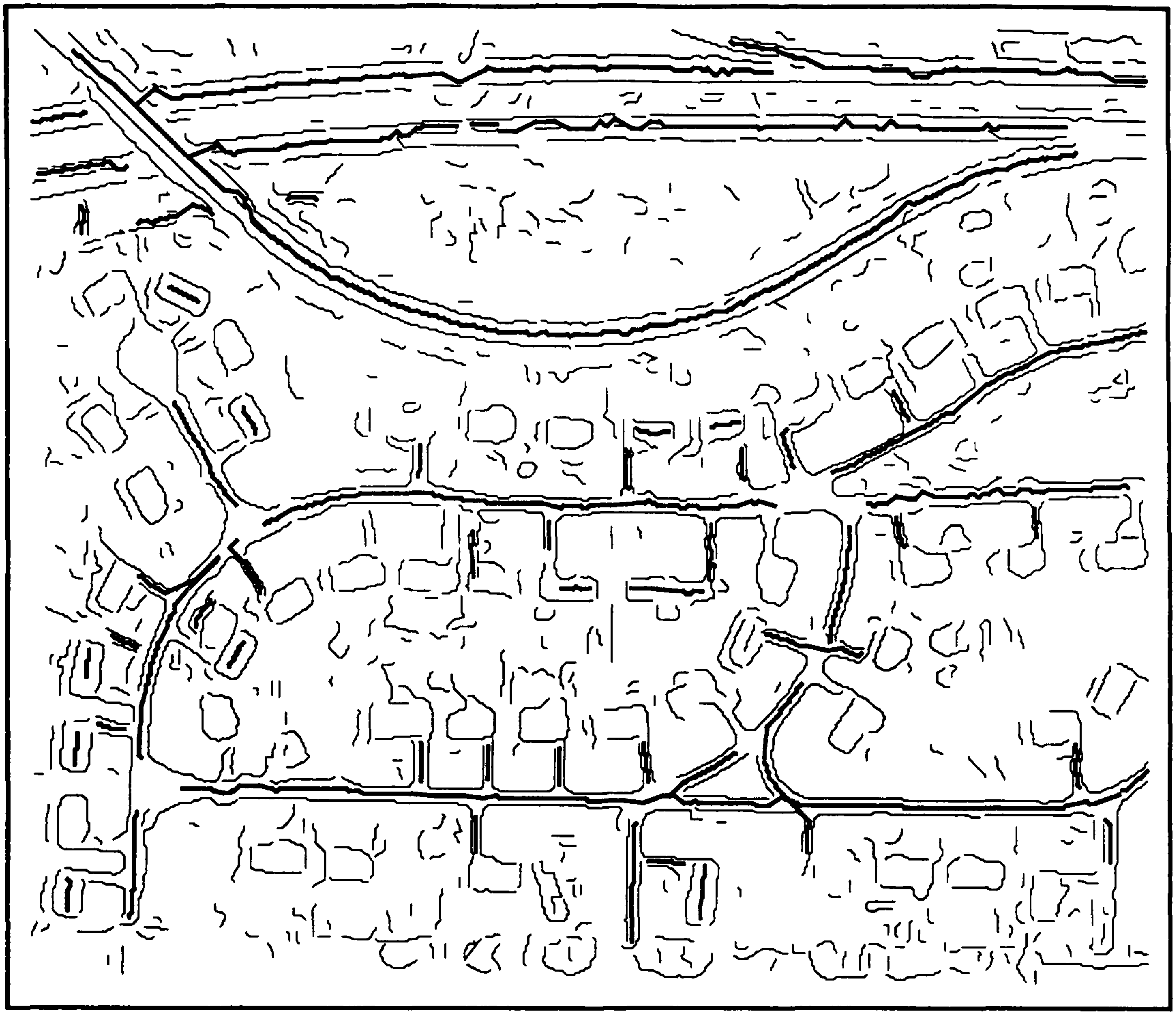

Figure 3-1: Road Seeds

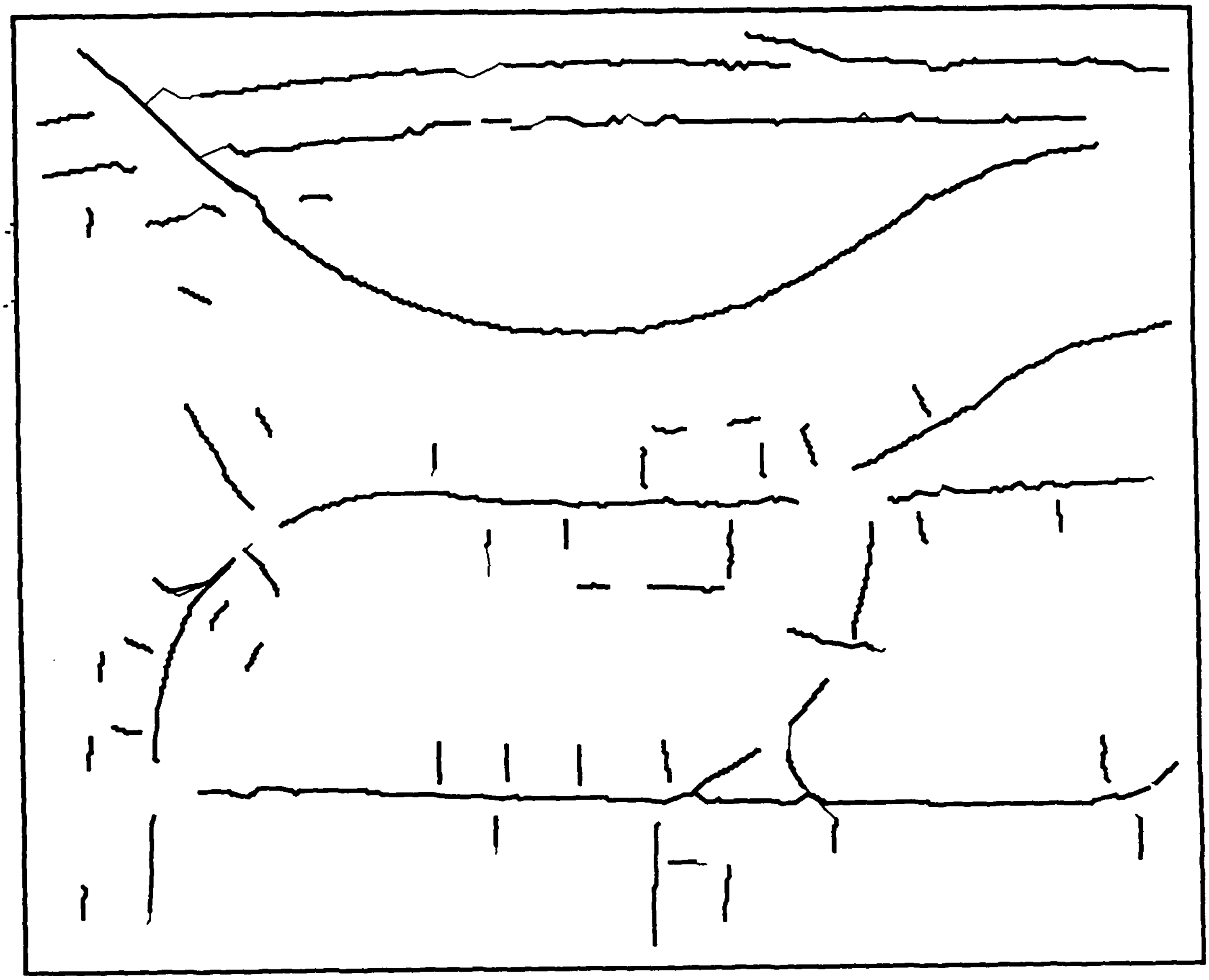

Figure 3-2: Detected Linking Errors (thin lines)

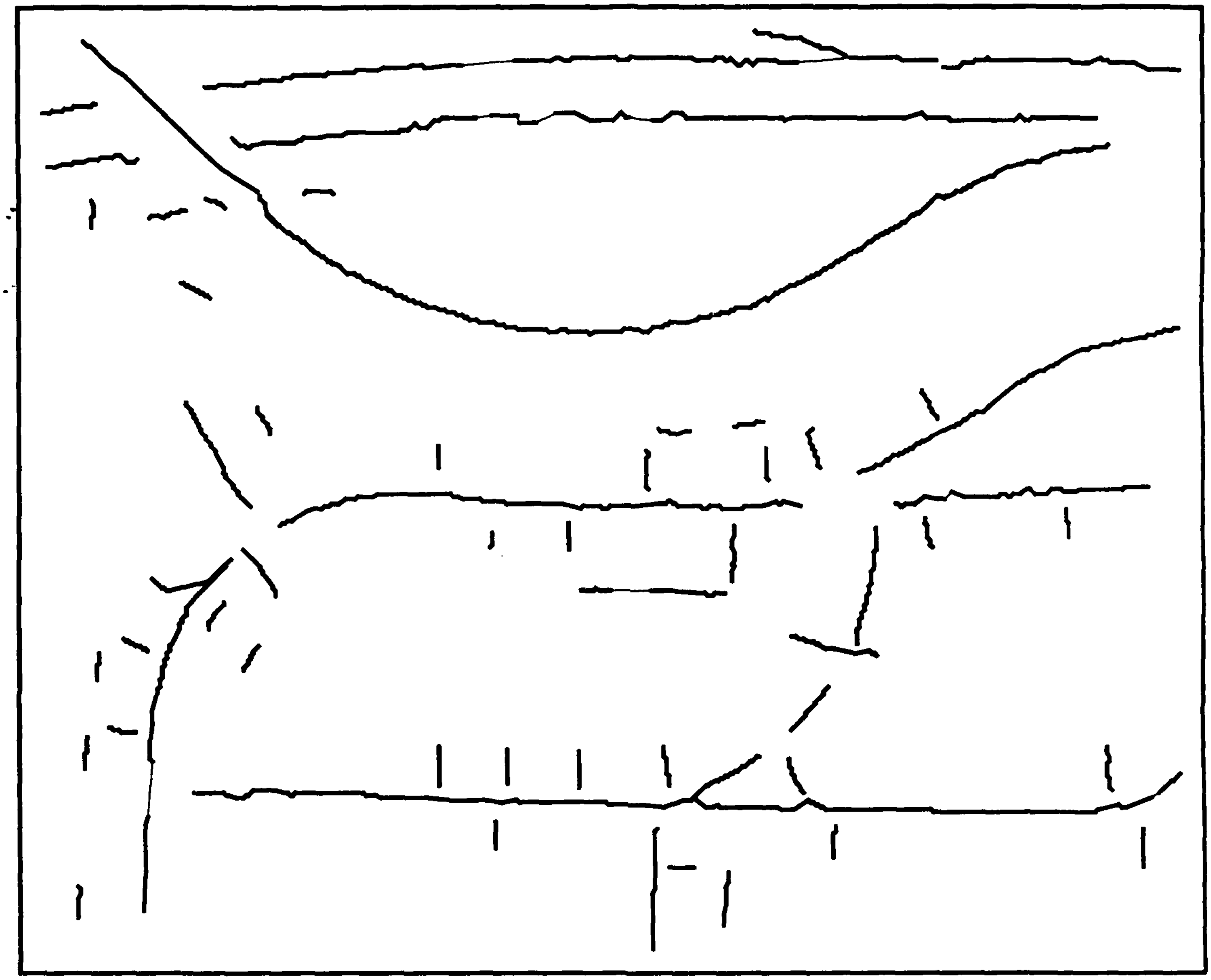

Figure 3-3: Post-Linking

3. **The road tracker** does most of the work. It is given road seeds and it tracks them, using two independent cooperating trackers. The basic road tracker ARF and its successor WOOF have been around for some time. In the last year the following modifications were made:
   a. Tracking was improved in two ways:
      i. Guessing ahead was changed so that only high confidence tracking steps will have influence on the projected path. This change was made possible by new scoring mechanisms developed last year. In these scores there is better correspondence between a step's score and its correctness. This change in the path projection eliminated many instances of the path being "locked" to the wrong direction because of a small number of erroneous steps.

      ii. A new computation of road-width at divergence points was introduced. Whereas previously the width from the better tracker was used to restart the failing tracker, now the spatial relation between the diverging paths is checked, and divergences that result from probable road width changes are detected and used to compute the new road widths. This change improves tracking especially when the edge-tracker loses one edge, as frequently happens on bridges and in road merges.

   b. The control strategy was updated so as to:
      i. Detect road splitting when trackers diverge and add the appropriate road seeds to the seed agenda.

      ii. Fire the road finder when the edge tracker looses an edge and cannot regain it within a reasonable number of tracking steps.

      iii. Fire the road finder when both trackers fail.

   All of these changes have a positive impact on the overall tracking quality.

Figure 3-4 shows road seeds found in a manually selected area of interest. Figure 3-5 shows all the roads tracked starting from these seeds. Figure 3-6 shows all the roads tracked without re-invoking the road finder and without returning to detected road splits.

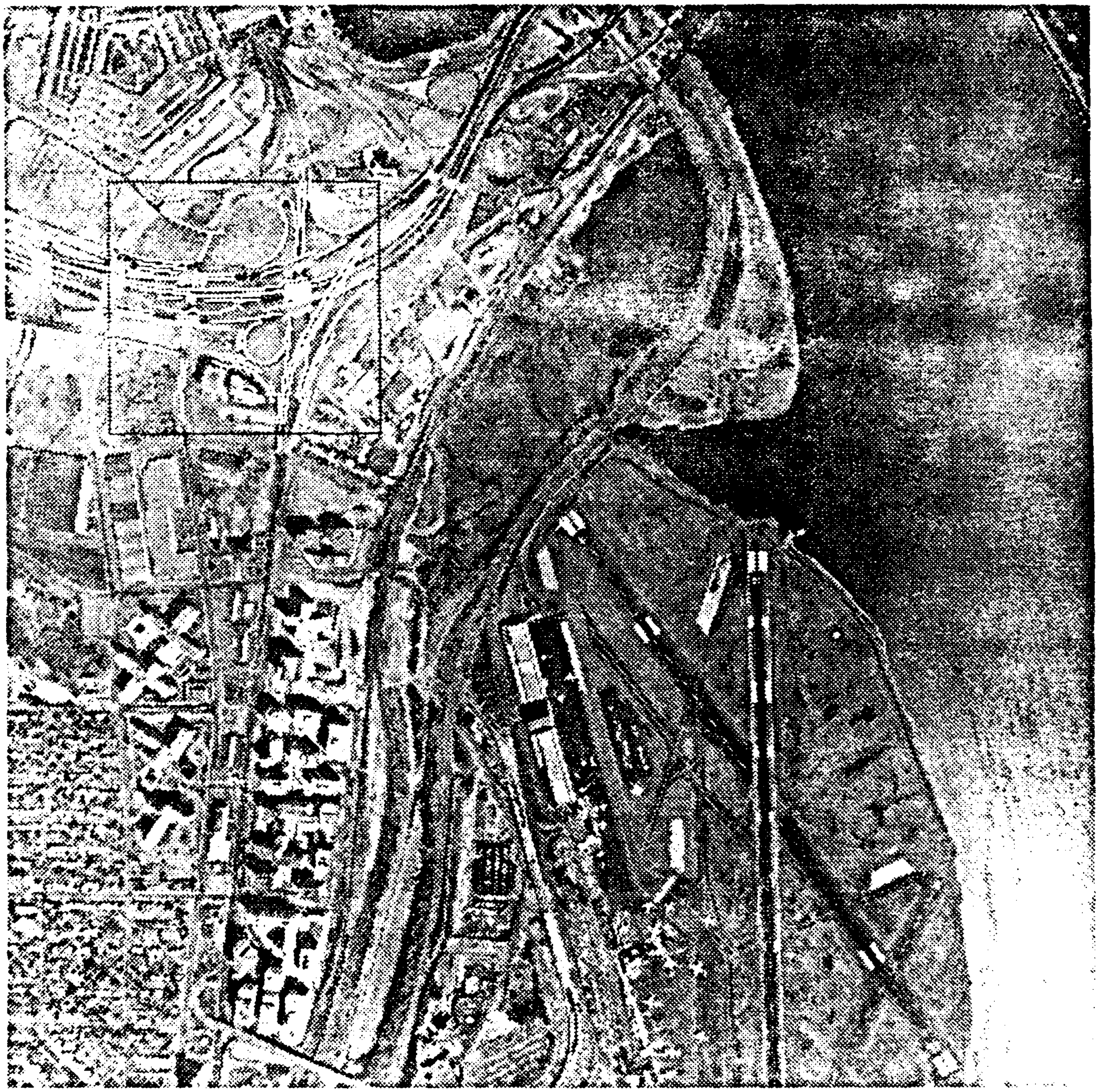

Figure 3-4: Road Seeds in Area of Interest

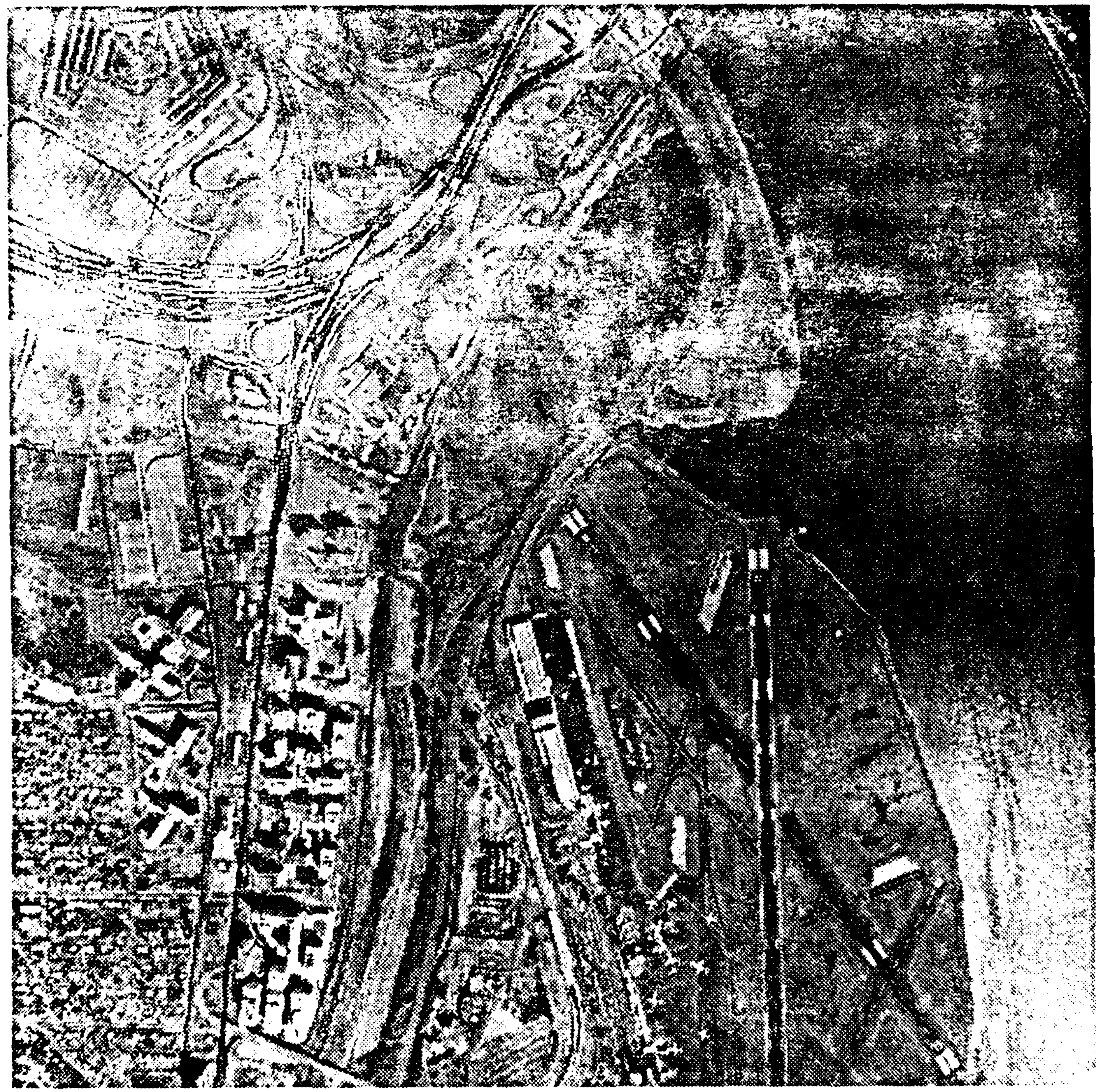

Figure 3-5: Tracked Roads

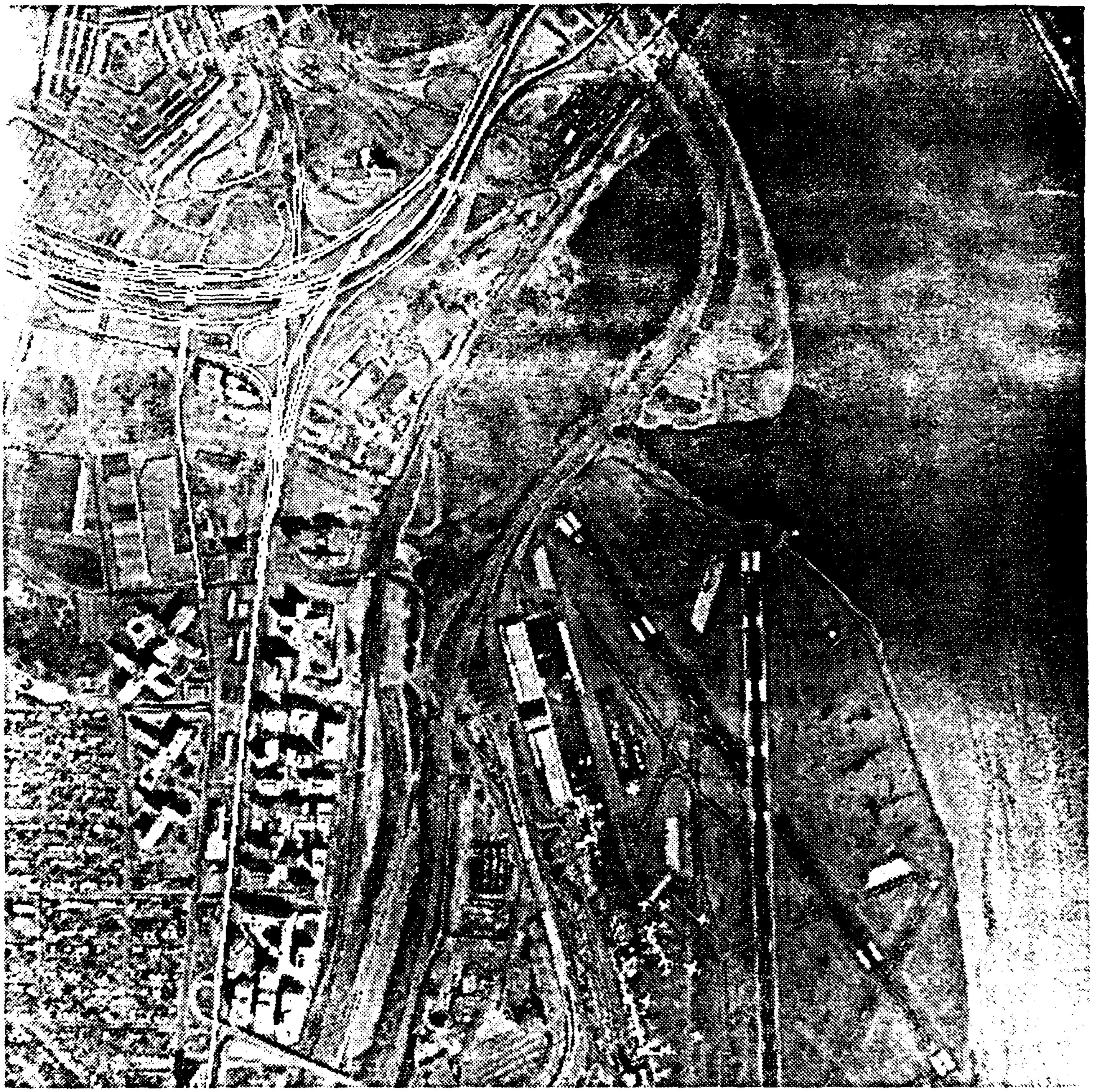

Figure 3-6: Tracking without splitting and re-finding

4. **The road network analyzer** takes many roads - output of the road tracker that was started with seeds from the road finder. It eliminates overlaps (usually resulting from merging) and marks intersections.

   The network analyzer works in two phases:

   a. All roads are painted into a bitmap, overlaps are detected, and roads are chopped accordingly. The program that does this is able to distinguish between real overlap and accidental overlap due to small random deviations of road paths.

   b. After roads are chopped a post-linking process is invoked. This process is similar to post linking in the road finding phase, but it also identifies intersections and labels them.

   Figure 3-7 shows a simple case of tracker output. Figure 3-8 shows the network analysis output.

3.1. Deficiencies and Future Work

Currently the road finder can not decide automatically whether roads are bright or dark, and if both kinds of roads coexist in the same section of the image the quality of the extracted network suffers. Its seems possible to correct this deficiency by using a better road model in the road finder.

Shadows are frequently interpreted as roads: the road finder finds seeds in them, and the road tracker happily extends those seeds.

The tracking phase introduces errors, mainly cases where road shoulders are tracked instead of the road itself. Such errors have a marked impact on the network analysis phase.

The road tracker produces a verbal description of the road followed and the features detected on it. This description is not integrated into the network construction yet.

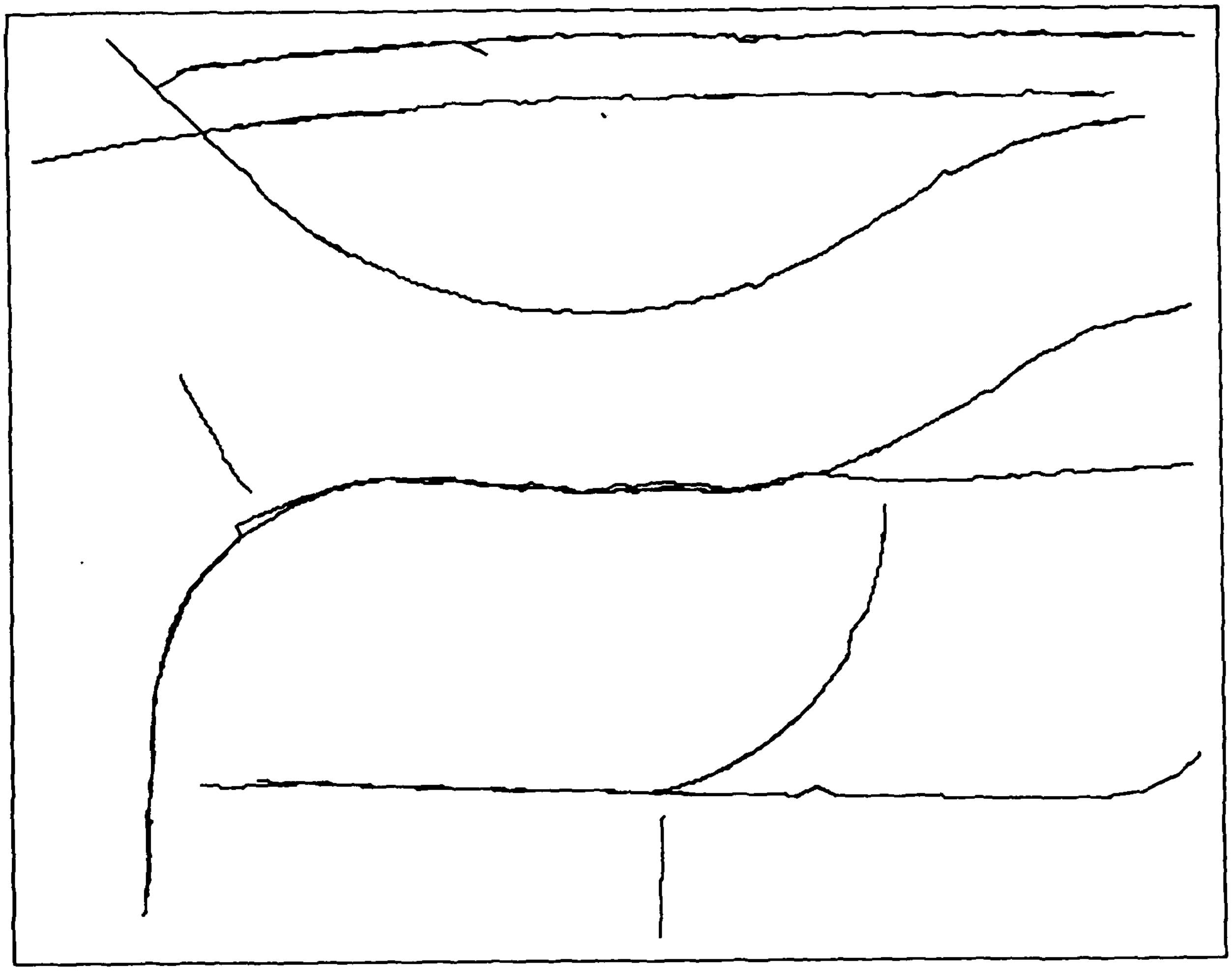

Figure 3-7: Tracker Output

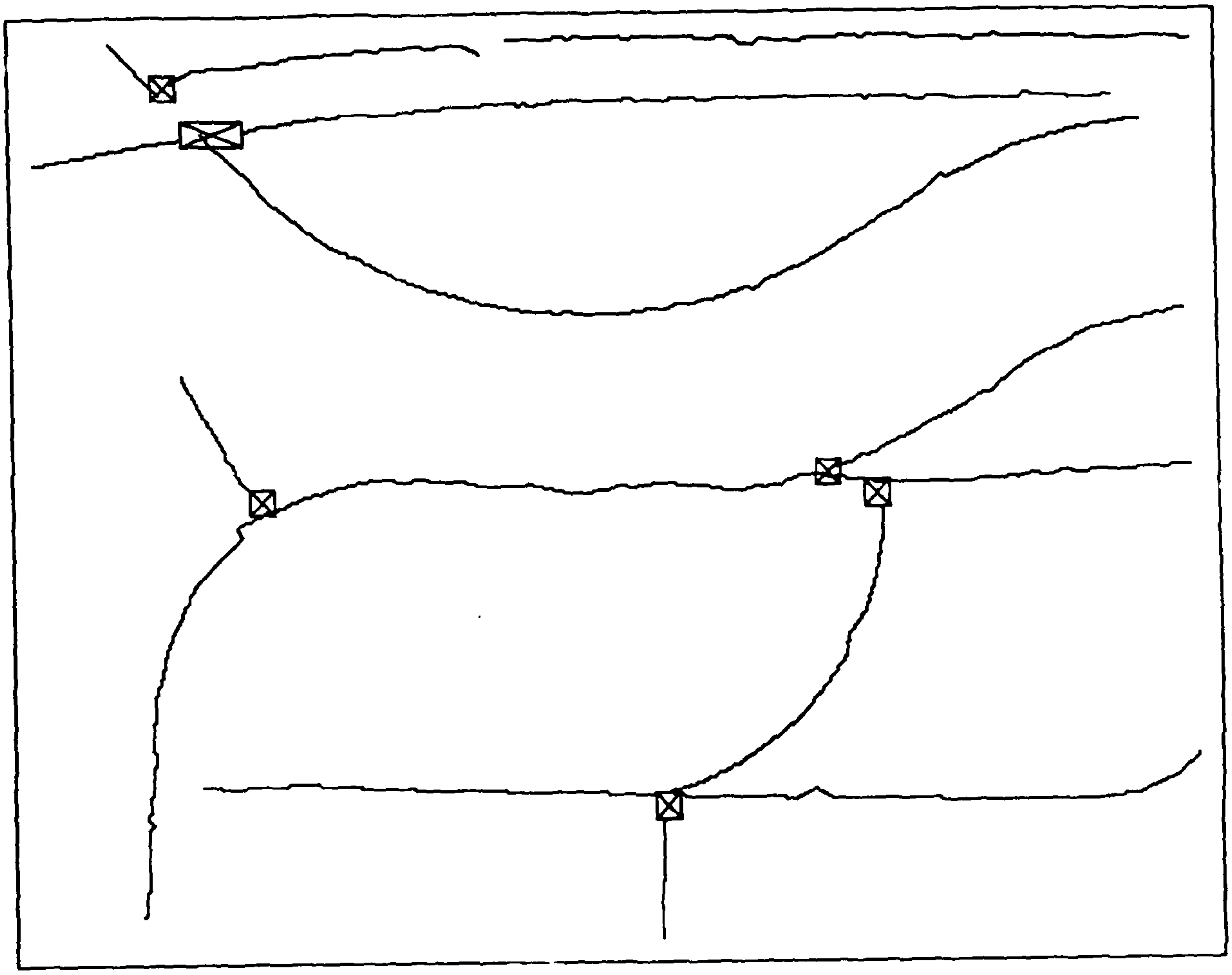

Figure 3-8: Constructed Network

4. Building Extraction in Builtup Areas

Buildings are another man-made feature that has importance in mapping applications, especially since buildings change quite frequently, and tracking all these changes manually is slow and expensive. Automatic detection of buildings requires delineation of objects in the image and selection of those objects that are actually buildings. Both these tasks are not trivial. First, the state of art in delineation of objects has left much for improvement. Techniques based on detecting patches of similar intensities (machine segmentation) are extremely sensitive to noise in the data - the samllest poor contrast area in a building may cause the building to be merged with its background. Edge detectors suffer from weakness both in sensitivity to noise and in poor performance near corners, which are after all the most distinctive feature of buildings. Second, distinguishing a building from other rectilinear features requires use of many different criteria. Some buildings may be recognized by their shadow, others by the 3-D projections of their walls, and yet others can be detected only by their similarity to other buildings in the scene.

It is our belief that no single method will prove sufficient for production quality building extractions. We are therefore designing a system in which machine segmentations, edge information and depth from stereo will all contribute to detection of buildings.

BABE (Builtup-Area Building Extraction) is the building extraction component based on intensity edges. It has four major phases:

1. **Finding possible building corners** After running a standard edge finding program we search for two kinds of corners:
   a. Corners formed by two edges that are relatively close and form an approximately right angle. This requires an efficient method to screen pairs of near edges, a task which is accomplished using the range search algorithm discussed in Section 5.

   b. Corners that are inside a single edge. Due to the way edge operators work, such corners are frequently rounded, and detecting them required the development of a special program, which we did. The program employs analysis of imperfect sequences, discussed in Section 5.

   Figure 4-1 shows the corners found in an image of LAX airport.

2. **Creating building hypotheses** Currently we are limited to box shaped buildings, but this restriction is expected to be removed in the near future. Building hypotheses are created by grouping edge segments that have common corners and a single rotation direction. Once such groups are created various boxes based on the group elements are created, and the best of those are selected as building hypotheses.

   Figure 4-2 shows the hypothesized boxes.

3. **Building verification** Boxes that have good edges, a uniform internal intensity and cast a shadow are good building candidates. We have developed a program that accepts a file of building hypotheses, guesses the illumination angle and the shadow intensity, and

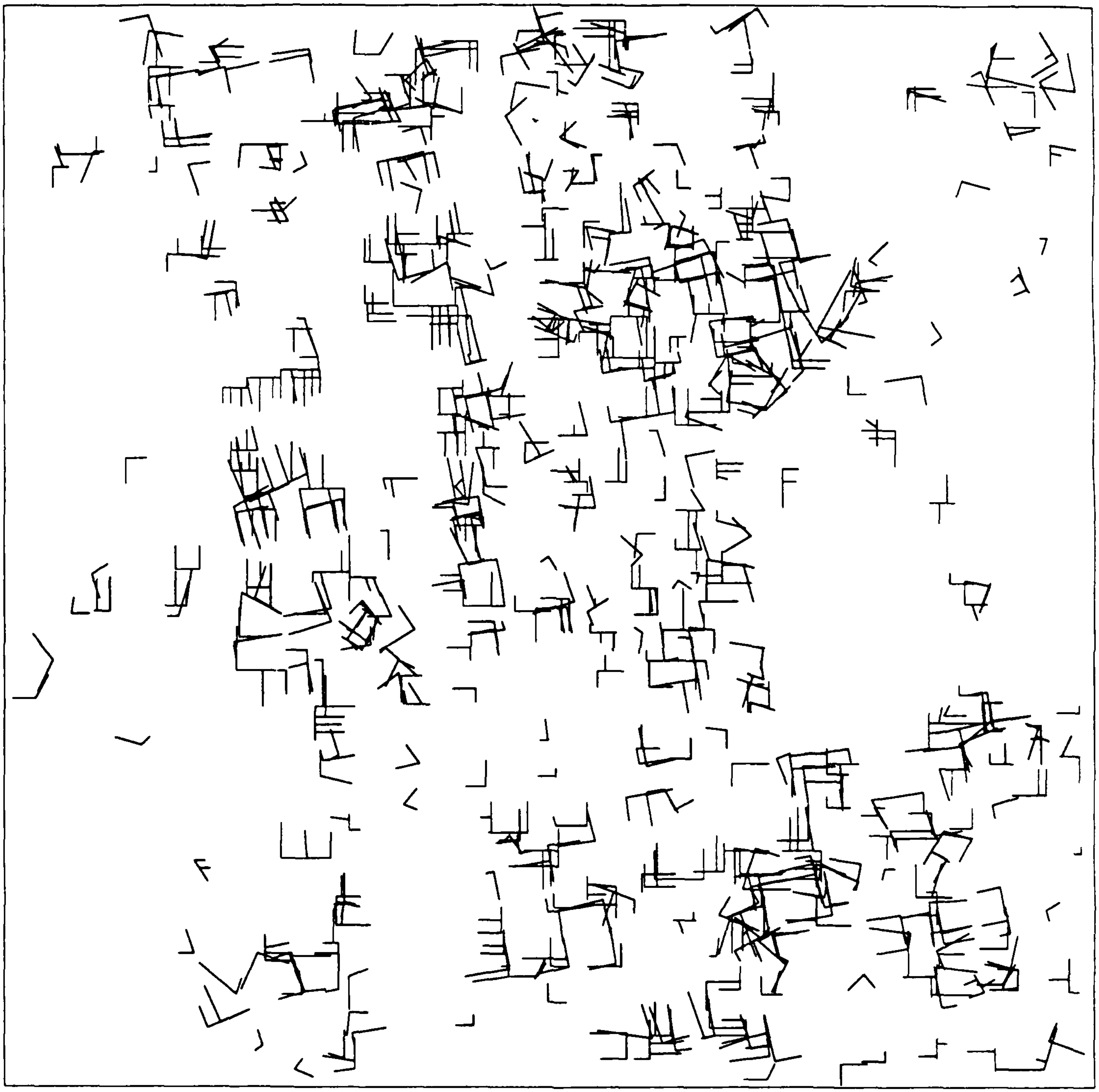

Figure 4-1: Corners

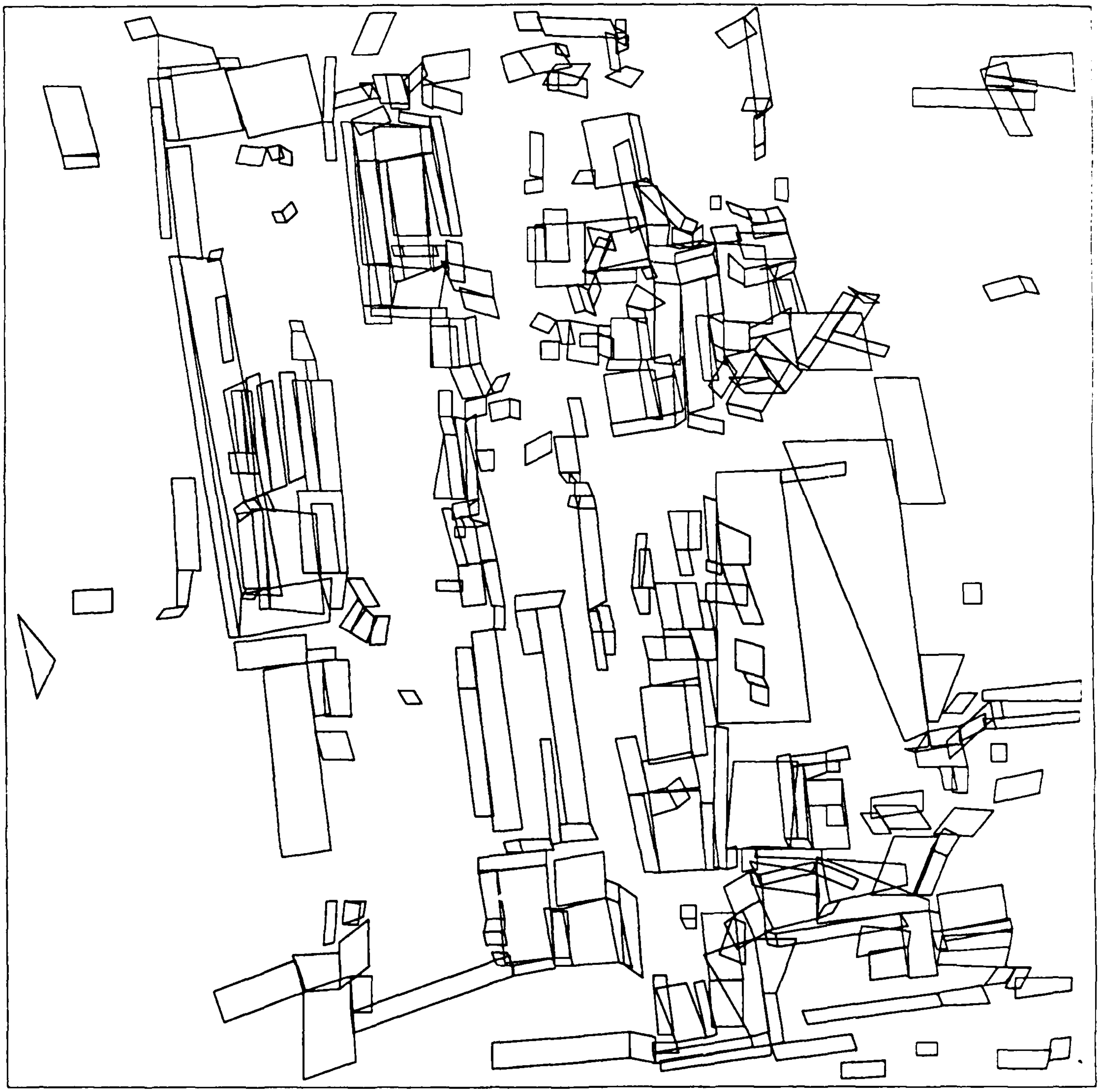

Figure 42: Edge-Based Building Hypotheses

selects box hypotheses that appear to be buildings. This process is very effective when buildings are box-shaped and not occluded, and performs reasonably well even when these constraints are somewhat relaxed.

Figure 4-3 shows the accepted hypotheses.

4. **Delineation improvement** Whereas the first three phases in BABE are content with having approximate building locations, the fourth phase aims to delineate the building precisely. This is not a simple task since many buildings have noisy edges and confusing shadows. However, we have an initial version of the delineation improving program, and the results indicate that the approach we have taken is correct.

   Figure 4-4 shows the results of delineation improvement. The white boxes are the accepted hypotheses, the black ones are the improvements. Note the long box in the upper right especially.

BABE has gone through three versions since work on it began last August. It currently has approximately 80% detection with less than 10% errors on the average in moderately complex images. Table 4-1 lists the performance of the various versions. As can be seen we have made some progress in eliminating false alarms while leaving the correct detection ration virtually stable.

It should be noted that although on the surface BABE seems to be similar to work done by Huertas and Nevatia in USC, our system does not assume the illumination angle is known, nor does it rely on any manually selected parameters. Furthermore, we are able to detect buildings whose roofs have shadow intensities, whereas the USC system does not seem to be able to do so.

Figures 4-3 4-5, 4-6, 4-7, 4-8 are the results on which the V3 column of table 4-1 are based.

4.1. Deficiencies and Future Work

It seems that the major source for missed buildings in BABE is the box generation step. The box generation algorithm is sensitive to noise, and fragmentation of buildings frequently causes them to be rejected in the verification phase. In addition to this there are some occurrences of false detections that seem to have very weak support from the image, indicating a weak spot in the verification process. These two issues are the first on the BABE agenda of future work.

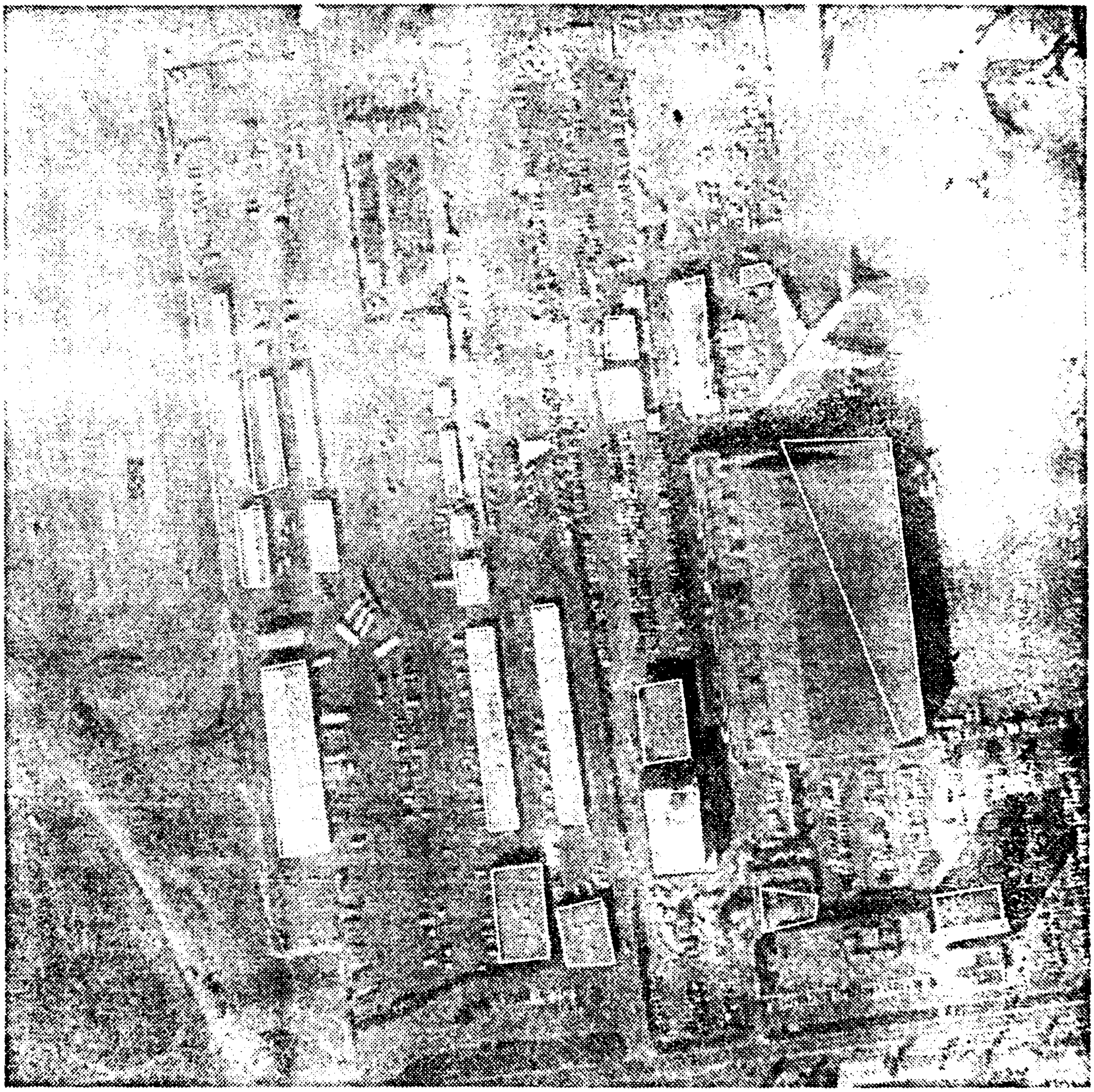

Figure 4-3: Buildings for image LAXW

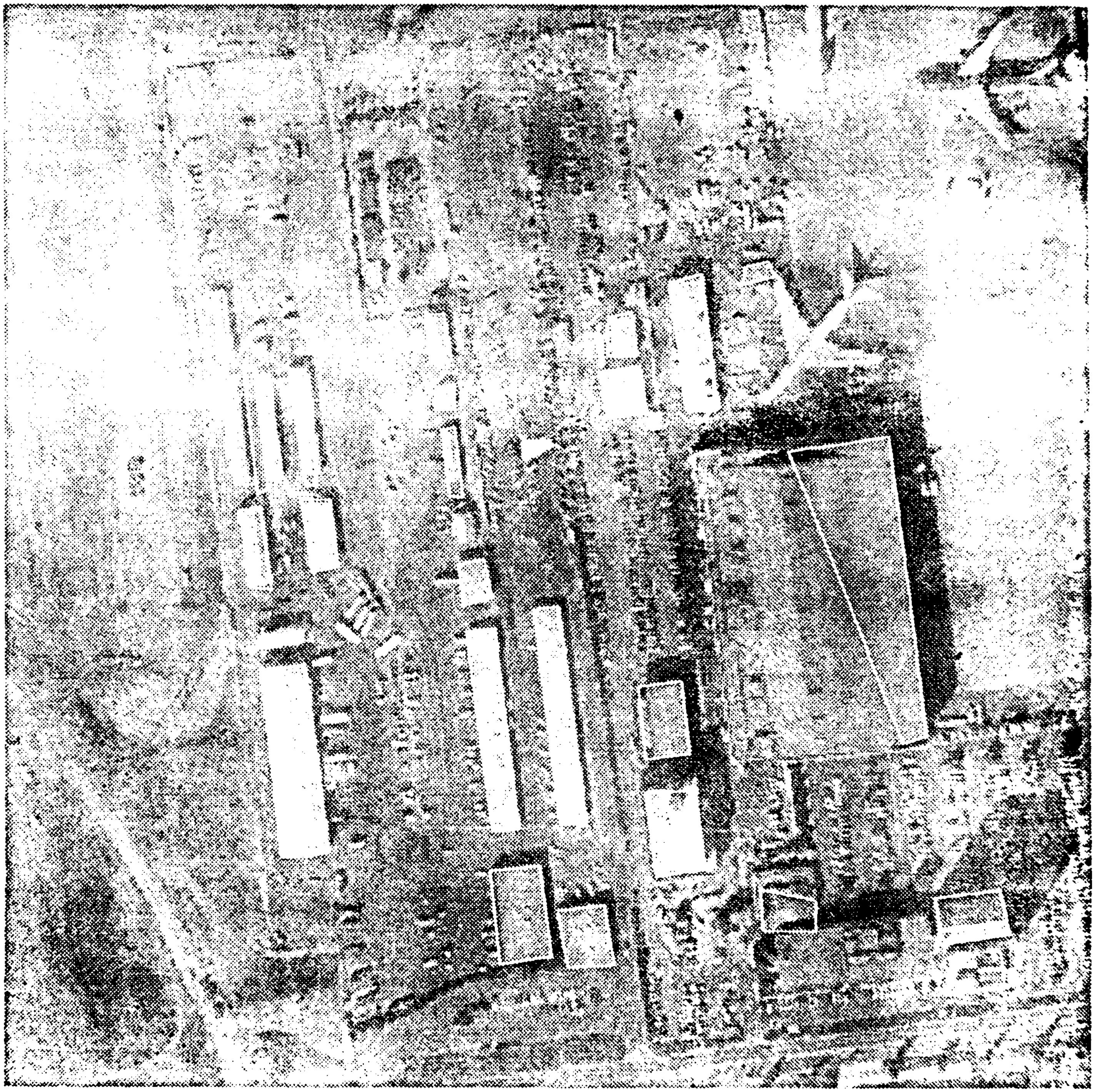

Figure 4-4: Improved Delineation

Performance	V1	V2	V3
SUBURB6:			
Number of buildings in scene:	52	52	52
Number of buildings correctly detected:	48	50	49
Number of buildings correctly delineated:	28	29	30
Number of buildings missed:	4	2	3
Number of incorrect detections:	2	1	2
detections /buildings =	92%	96%	94%
delineations / buildings =	53%	55%	57%
incorrect / buildings =	7%	2%	4%
correct / incorrect =	24.0	52.0	24.5
LAX80W			
Number of buildings in scene:	25	25	25
Number of buildings correctly detected:	22	21	20
Number of buildings correctly delineated:	13	18	9
Number of buildings missed:	3	4	3
Number of incorrect detections:	10	8	1
Number of alligned building fragments:	4	6	5
detections /buildings =	88%	84%	80%
delineations / buildings =	52%	64%	36%
incorrect / buildings =	40%	32%	4%
correct / incorrect =	2.2	2.6	20.0
LAX80E			
Number of buildings in scene:	26	26	26
Number of buildings correctly detected:	18	22	19
Number of buildings correctly delineated:	13	14	10
Number of buildings missed:	8	4	3
Number of incorrect detections:	8	14	6
Number of alligned building fragments:	4	5	5
detections /buildings =	69%	84%	74%
delineations / buildings =	50%	54%	36%
incorrect / buildings =	30%	54%	22%
correct / incorrect =	2.25	1.56	3.2
DC37405			
Number of buildings in scene:	88	88	88
Number of buildings correctly detected:	53	59	65
Number of buildings correctly delineated:	31	36	25
Number of buildings missed:	32	26	16
Number of incorrect detections:	17	9	9
Number of alligned building fragments:	2	2	8
detections /buildings =	60%	67%	74%
delineations / buildings =	35%	40%	30%
incorrect / buildings =	19%	9%	8%
correct / incorrect =	3.11	6.5	8.0
DC38008			
Number of buildings in scene:	36	36	36
Number of buildings correctly detected:	14	15	14
Number of buildings correctly delineated:	7	9	5
Number of buildings missed:	22	22	16
Number of incorrect detections:	20	20	6
Number of alligned building fragments:	18	12	21
detections /buildings =	38%	41%	38%
delineations / buildings =	19%	24%	14%
incorrect / buildings =	61%	61%	17%
correct / incorrect =	0.7	0.7	2.3

Table 4-1: BABE Performance Summary

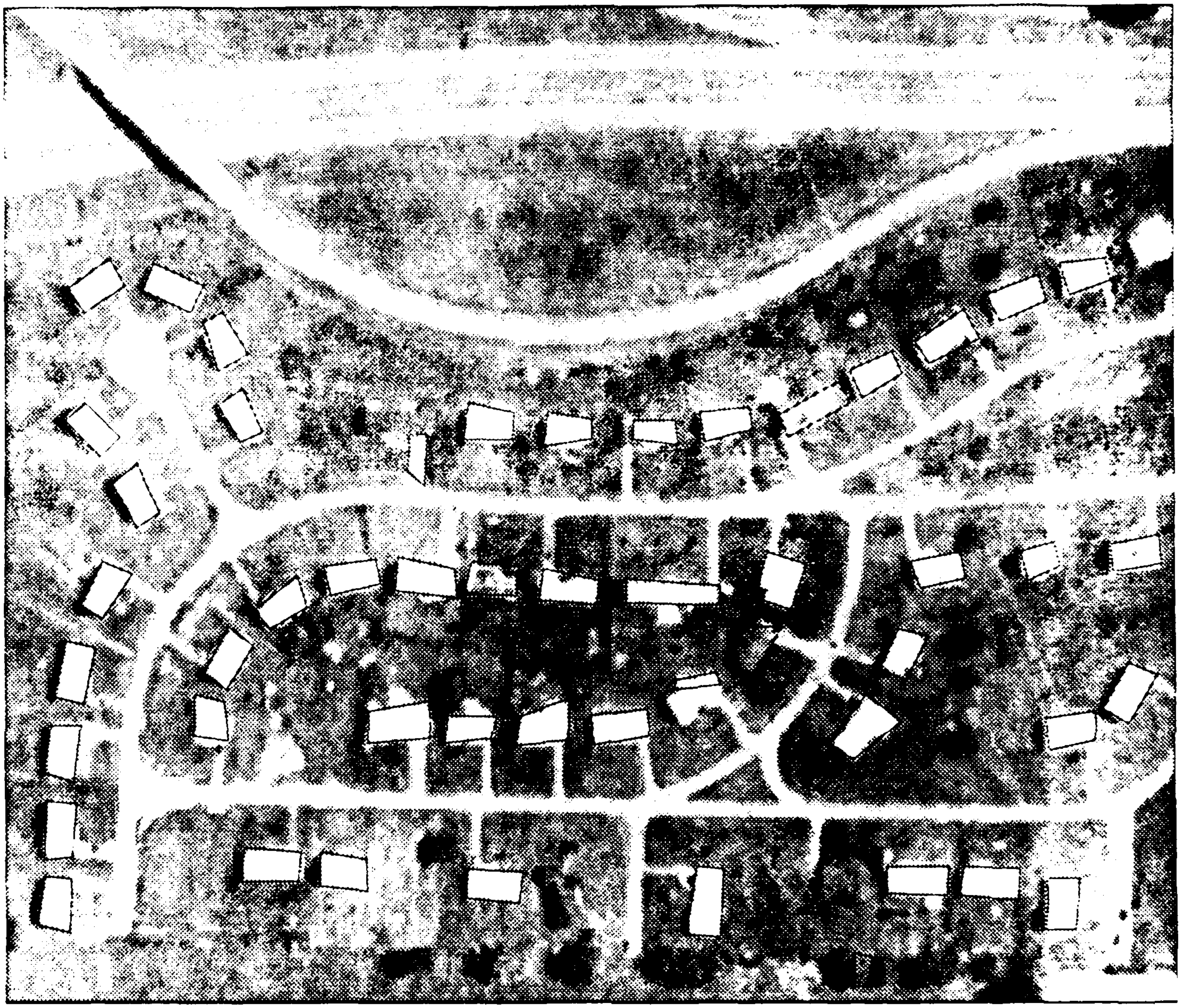

Figure 4-5: Buildings for image SUBURB6

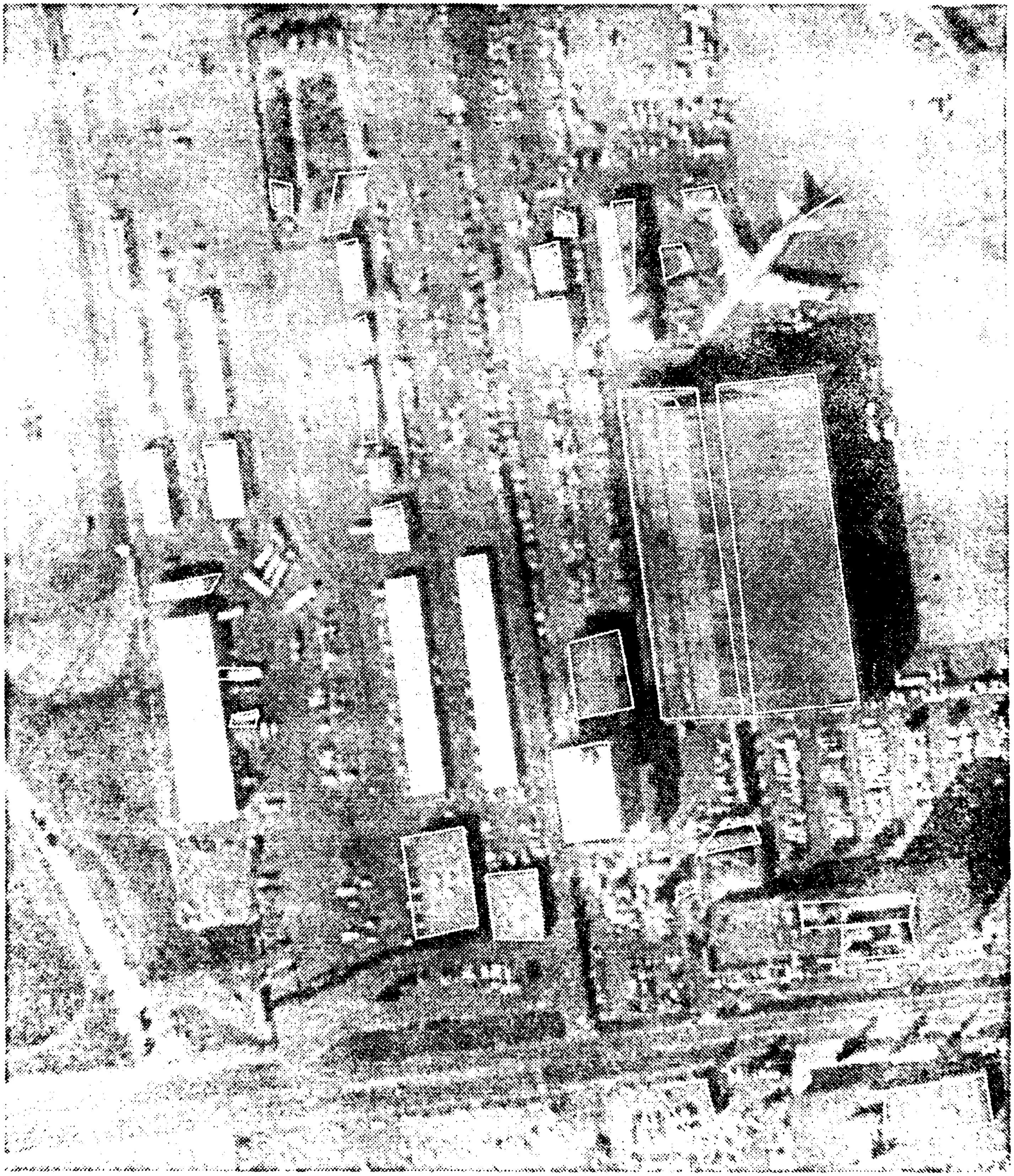

Figure 4-6: Bu ldings for mage LAXE

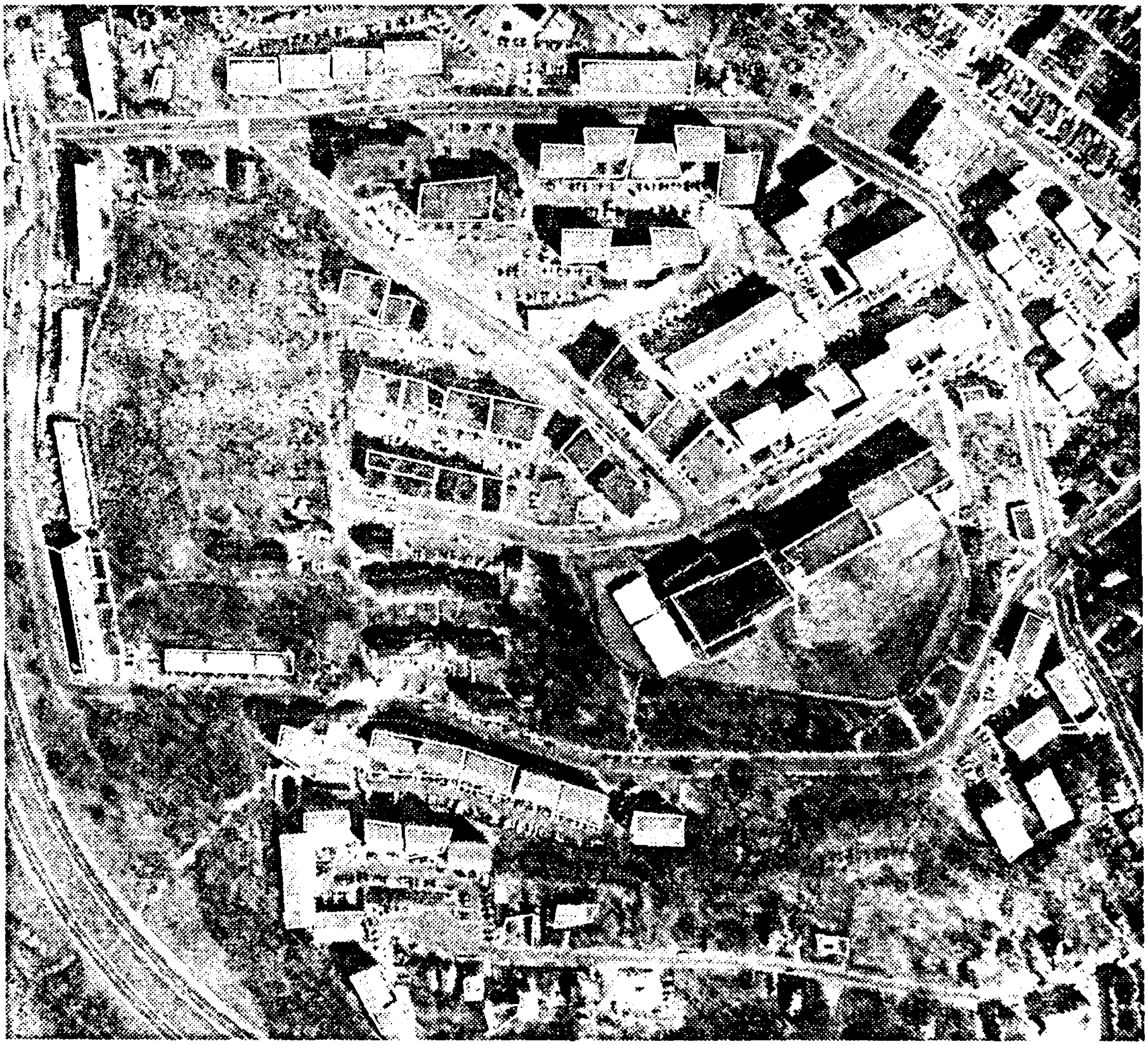

Figure 4-7: Buildings for image DC37405

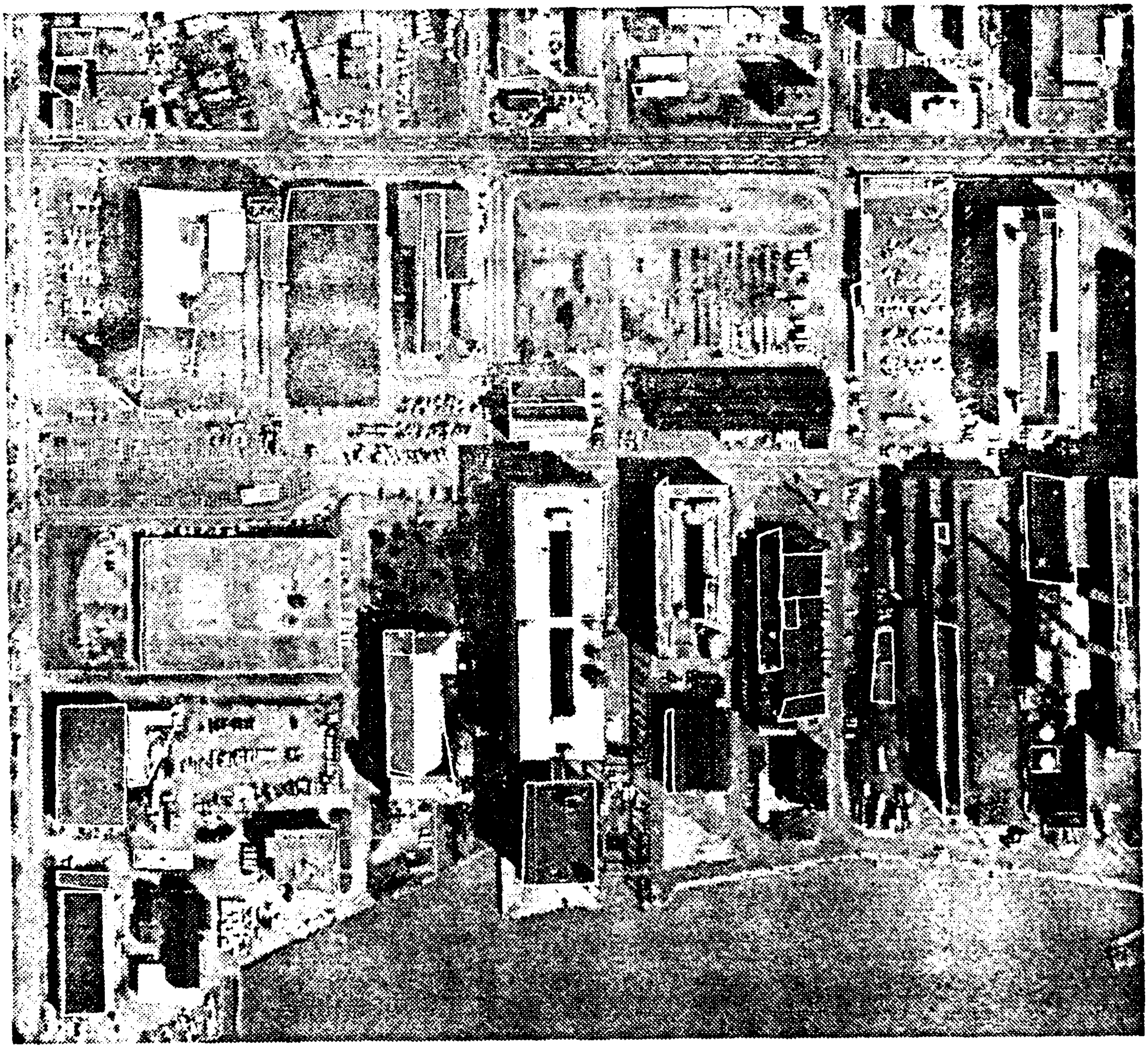

Figure 4-8: Buildings for image DC38008

5. Tools

In the course of this year we have implemented a number of general purpose tools that are used in several applications:

1. **An equivalence class abstract data type** This is an implementation of a standard algorithm, but it has come in handy in several applications: in post-linking and intersection finding in RNET, and in structure generation in BABE. Our implementation includes a home-made memory management scheme that relieves users of the need to declare and initialize the data structure, making its use especially easy.

2. **A multiple heap abstract data type** This too is an implementation of a standard algorithm. It is used in several places in RNET and in BABE, and employs the same memory management scheme.

3. **A range search algorithm** This algorithm permits retrieval of pairs of line segments that are spatially close without having to go through an exhaustive enumeration of all line segments in question. Timing tests show that execution time is speeded up by at least an order of magnitude, enabling us to process large images on conventional hardware where others have to limit themselves to small images and use expensive machines.

   The range search module is used in RNET for post-linking and network construction, in BABE for finding corners and matching hypotheses to ground truth or to results from different resolutions, and is also used in a new feature based stereo program that we are developing.

   The range search algorithm and its implementation are our own development, and are especially suited for processing related to discrete raster images.

4. **Imperfect Sequence Analysis** This is an algorithm that detects linear objects in presence of measurement errors and noise. It is used extensively throughout RNET (in road finding, feature detection during tracking, control of the cooperating processes, and detection of road overlaps in the network construction phase), and is the basic building block used for detecting intra-edge corners in BABE. It is also used in BABE's delineation improvement phase and in two shadow analysis programs that we have developed.

   The sequence analysis algorithm is a heuristic program that balances local and global interpretations of binary test results. In contrast with other global optimization programs, imperfect sequence analysis is surprisingly fast without any observed performance compromise.

5. **Line hypothesis evaluation** In BABE one frequently has incomplete corner-edge information on a building. In such cases it is necessary to hypothesize the location of undetected building sides and corners, and to examine the image intensity for support of the hypothesis. Since we are not limiting ourselves to bright buildings on a dark

background, it was necessary to develop an edge evaluation method that is not based on contrast alone. Our success in finding such a method enabled us to detect buildings with dark roofs even when the contrast between the roof and the shadow is poor.

Our line evaluation procedure is based on detecting places that have a high gradient in relation to their neighborhood, and analyzing the distribution of those places around the hypothesized line. The two versions that we have already implemented perform reasonably well, and have given us ideas how to achieve further improvement.

5.1. Year Two Research Agenda

During the second year of this contract we plan to solidify the initial results achieved in road tracking and building extraction. As previously noted there are several deficiencies in the current BABE system dealing with the aggregation of corner-line structures into boxes. We believe that there are several improvements that can be made which take into account local image intensity profiles to attempt to measure the amount of noise in the line and corner approximations. In addition to these evolutionary improvements we plan to begin work in the new areas of shadow analysis and stereo matching.

- Shadow analysis has been largely overlooked in the computer vision literature as a method for verifing and predicting the location of man-made structures in aerial imagery. This is interesting because it is clear that photo-interpreters make use of shadow information when presented with monocular imagery. We plan to explore ways to automatically detect and delineate shadows in aerial imagery and apply such delineations to building analysis.

- Stereo analysis provides a direct measurement of the three-dimensional structure of an aerial imagery. However most stereo matching algorithms have great difficulty in handling large depth discontinuities found in urban scenes. Feature based techniques that provide disparity estimates only at matchable points show promise in providing a more robust estimate in complex scenes. However, many open problems remain in accurate scene registration and in methods for interpolation between sparse disparity points to achieve a dense estimate of surface elevation.

6. Publications, Reports, Presentations

During the first year of this contract we have begun to publish our research results in the scientific literature. The following details a list of talks and publications. Copies of published papers are appended to this final report.

6.1. Publications

1. "Building knowledge-based systems for detecting man-made structures from remotely sensed imagery", McKeown, D. M., in *Philosophical Transactions of the Royal Society London*, pages 423-435, March, 1988, Volume A324.

2. "Cooperative Methods for Road Tracking in Aerial Imagery", McKeown, D.M. and Denlinger, J. L., in *Proceedings IEEE Computer Vision and Pattern Recognition Conference*, pages 662-672, Ann Arbor, MI, June 1988.

3. "A Discrete Scale-Space Representation", Aviad, Z., in *Proc. 1st International Conference on Computer Vision*, London, England, June 1987.

4. "Road Finding for Road Network Extraction", in *Proceedings IEEE Computer Vision and Pattern Recognition Conference*, Aviad, Z. and P. D. Carnine, pages 814-819, Ann Arbor, MI, June 1988.

6.2. Talks and Presentations

David M. McKeown, Jr

"Cooperative Methods For Road Tracking In Aerial Imagery"
IEEE Conference on Computer Vision and Pattern Recognition
Ann Arbor, Michigan, June 9, 1988.

"Cooperative Methods For Road Tracking In Aerial Imagery"
DARPA Image Understanding Workshop,
Boston, Massachusetts, April 8, 1988.

"Automated Feature Extraction From Aerial Imagery",
EXRAND Committee Meeting,
Washington, D.C., February, 9, 1988.

"Cooperative Feature Extraction"
Sixteenth IEEE Workshop on Applied Imagery Pattern Recognition,
Washington, D.C. October 29, 1987.

Tutorial on "Spatial Interpretation of Aerial Imagery"
at *IEEE Workshop on Applied Imagery Pattern Recognition*,
Washington, D.C., October 28, 1987.

Aviad Zlotnick

"Road Finding for Road-Network Extraction"
IEEE Conference on Computer Vision and Pattern Recognition
Ann Arbor, Michigan, June 7, 1988.

7. References

[1] Avron Barr and Edward A. Feigenbaum.
The Handbook of Artificial Intelligence.
William Kaufmann, Inc., Los Altos, CA, 1981.

[2] McKeown, D.M.
MAPS: The Organization of a Spatial Database System Using Imagery, Terrain, and Map Data.
In *Proceedings: DARPA Image Understanding Workshop*, pages 105-127. June, 1983.
Also available as Technical Report CMU-CS-83-136.

[3] McKeown, D.M., and McDermott, J.
Toward Expert Systems for Photo Interpretation.
In *IEEE Trends and Applications '83*, pages 33-39. May, 1983.

[4] McKeown, D.M.,
Digital Cartography and Photo Interpretation from a Database Viewpoint.
In Gargarin, G. and Golembe, E. (editor), *New Applications of Databases*, pages 19-42. Academic Press, New York, N. Y., 1984.

[5] McKeown, D. M.
Knowledge-Based Aerial Photo Interpretation.
Photogrammetria, Journal of the International Society for Photogrammetry and Remote Sensing 39:91-123, 1984.
Special Issue on Pattern Recognition.

[6] McKeown, D.M., Harvey, W.A. and McDermott, J.
Rule Based Interpretation of Aerial Imagery.
IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-7(5):570-585, September, 1985.

[7] McKeown, D.M.
The Role of Artificial Intelligence in the Integration of Remotely Sensed Data with Geographic Information Systems.
IEEE Transactions on Geoscience and Remote Sensing GE-25(3):330-348, May, 1987.

[8] McKeown, D.M., and Harvey, W.A.
Automating Knowledge Acquisition For Aerial Image Interpretation.
Technical Report CMU-CS-87-102, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA. 15213, 1987.
Extended version in CVGIP Volume 46 No.1 April 1989.